DOCUMENT RESUME

ED 452 638 EC 308 395

AUTHOR McConnell, Scott; McEvoy, Mary; Carta, Judith J.; Greenwood, Charles R.; Kaminski, Ruth; Good, Roland H., III; Shinn, Mark

TITLE Research and Development of Individual Growth and Development Indicators for Children between Birth to Age Eight. Technical Report #4.

INSTITUTION Minnesota Univ., Minneapolis. Early Childhood Research Inst. on Measuring Growth and Development.

SPONS AGENCY Special Education Programs (ED/OSERS), Washington, DC. Early Education Program for Children with Disabilities.

PUB DATE 1998-06-00

NOTE 39p.; With assistance from James Ysseldyke and Paula Goldberg. For related technical reports, see EC 308 392-394 and EC 396-398.

CONTRACT H02450010

AVAILABLE FROM Early Childhood Research Institute on Measuring Growth and Development, University of Minnesota, 202 Pattee Hall, 150 Pillsbury Dr., S.E., Minneapolis, MN 55455. Tel: 612-624-8020; Fax: 612-625-2093; e-mail: pries005@umn.edu.

PUB TYPE Reports - Evaluative (142)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Child Development; Developmental Stages; *Disabilities; Early Childhood Education; *Early Intervention; Evaluation Methods; *Individual Development; Preschool Children; Primary Education; *Standards; Student Educational Objectives; *Student Evaluation; Young Children

ABSTRACT

This document provides an overview of the rationale for, and characteristics of, individual growth and development indicators (IGDIs) for children birth to age 8 and their families. Development of such indicators is part of a 5-year project by the Early Childhood Research Institute Measuring Growth and Development to conduct research on, develop, and disseminate a comprehensive program performance measurement system for early intervention, preschool, and primary-grade programs serving children with disabilities and their families. The IGDIs are intended to describe children's growth and development over time, and thus to indicate when this progress is on-track toward some desired outcome, or when different or more intensive intervention is needed. This paper looks at the purpose and conceptual foundation for IGDI assessment, describes essential and desirable features of these measures, offers examples of possible IGDIs, and reports findings from the initial research on these measures. (Contains 42 references.) (DB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 452 638

EARLY CHILDHOOD RESEARCH INSTITUTE
ON

MEASURING GROWTH & DEVELOPMENT

TECHNICAL REPORT #4

RESEARCH AND DEVELOPMENT OF INDIVIDUAL GROWTH AND DEVELOPMENT INDICATORS FOR CHILDREN BETWEEN BIRTH TO AGE EIGHT

EC 308 395

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

The work reported here was supported by Grant No. H02450010, funded by the Early Education Programs for Children with Disabilities (EEPCD), Office of Special Education and Rehabilitation Services (OSERS), U.S. Department of Education. The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

Feel free to copy or distribute any part of this report. Please give credit to the Early Childhood Research Institute on Measuring Growth and Development.

For more information, contact the Early Childhood Research Institute on Measuring Growth and Development, University of Minnesota, 202 Pattee Hall, 150 Pillsbury Drive S.E., Minneapolis, MN 55455 (Phone: 612-624-8020, Fax: 612-625-2093, Email:pries005@umn.edu).

BEST COPY AVAILABLE

EARLY CHILDHOOD RESEARCH INSTITUTE
ON

MEASURING GROWTH & DEVELOPMENT

TECHNICAL REPORT #4

RESEARCH AND DEVELOPMENT OF INDIVIDUAL GROWTH AND DEVELOPMENT INDICATORS FOR CHILDREN BETWEEN BIRTH TO AGE EIGHT

APRIL 1998

A Collaborative Effort

Scott McConnell & Mary McEvoy
Center on Early Education and Development
Institute on Community Integration
University of Minnesota

Judith J. Carta & Charles R. Greenwood
Juniper Gardens Children's Project
University of Kansas

Ruth Kaminski, Roland H. Good III, & Mark Shinn
Center on Human Development
University of Oregon

with assistance from

James Ysseldyke
National Center on Educational Outcomes
University of Minnesota

Paula Goldberg
PACER, Inc.
Minneapolis, Minnesota

Table of Contents

Abstract 1

Introduction 1

Purpose and Conceptual Foundation of Individual Growth and Development Indicators 2

Criteria or Desirable Features for Individual Growth and Development Indicators 5

Research Plans and Strategies for Formative Process of IGDI Development 9

Special Issues, Examples, and Results to Date, Individual Growth and Development for Three Age Groups 12

- Special Requirements and Considerations for Assessing Growth and Development for Infants and Toddlers, Preschoolers, and Early Elementary Students 13
- Application Example: Developing Individual Growth and Development Indicators for "Using Gestures, Words, or Sentences to Communicate Wants and Needs or Express Meaning to Others" 17

Using Individual Growth and Development Indicators in A Problem-Solving Assessment Model 28

Conclusion 30

Reference 31

Abstract

The Early Childhood Research Institute on Measuring Growth and Development (ECRI-MGD) was funded by the Office of Special Education Programs, U.S. Department of Education to conduct research, development, and dissemination to produce a comprehensive program performance measurement system for early intervention, preschool, and primary-grade programs serving children with disabilities from birth to eight years of age, and their families. One significant component of ECRI-MGD's work is the development and evaluation of individual growth and development indicators. These IGDIs are intended to describe children's growth and development over time, and thus to indicate when this progression is on-track toward some desired outcome, or when different or more intensive intervention is needed (see ECRI-MGD Technical Report 5). Thus, individual growth and development indicators constitute one significant portion of the emerging comprehensive program performance measurement system for early intervention, preschool, and primary-grade programs serving children with disabilities from birth to eight years of age, and their families. The purpose of this paper is to provide a broad overview of the rationale for, and characteristics of, individual growth and development indicators for children birth to age eight and their families. In particular, we will provide an overview of the purpose and conceptual foundations for individual growth and development indicator assessment; describe essential and desirable features of these measures, as well as the overall research and development process being followed by ECRI-MGD investigators to develop, select, and validate individual IGDIs; and provide examples of possible IGDIs and overview findings from the initial research on these measures.

Introduction

The Early Childhood Research Institute on Measuring Growth and Development (ECRI-MGD) was funded by the Office of Special Education Programs, U.S. Department of Education to research, develop, and disseminate a comprehensive program performance measurement system for early intervention, preschool, and primary-grade programs serving children with disabilities from birth to eight years of age and their families. The work of this institute follows increased attention to accountability and high-stakes assessment of educational and other progress in American schools (see ECRI-MGD Technical Report 1), and reflects a strong theoretical and conceptual commitment to frequent, repeated, dynamic and idiographic assessment that describes growth and development across time and directly assists in the design and evaluation of intervention efforts to support developmental progress (see ECRI-MGD Technical Report 6, as well as Deno, 1985; Good & Kaminski, 1996).

One significant component of ECRI-MGD's work is the development and evaluation of individual growth and development indicators. These individual growth and development indicators (or IGDIs) relate to each identified general outcome for young children and families (see ECRI-MGD Technical Reports 2 and 3), and provide ways for parents and practitioners to gather information that describes changes over time for individual children or for groups of children in classrooms, programs, districts, or states. These IGDIs describe children's growth and development over time, and thus can be used (along with formal decision rules) to indicate when this progression is on-track toward some desired outcome or when different or more intensive intervention is needed (see ECRI-MGD Technical Report 5). Thus, individual growth and development indicators constitute one significant portion of the emerging comprehensive program performance measurement system for early intervention, preschool, and primary-grade programs serving children with disabilities from birth to eight years of age, and their families (see ECRI-MGD Technical Report 6).

The purpose of this paper is to provide a broad overview of the rationale for, and characteristics of, individual growth and development indicators for children birth to age eight and their families. In particular, we will (a) provide an overview of the purpose and conceptual foundations for individual growth and development indicator assessment; (b) describe essential and desirable features of these measures; (c) describe the overall research and development process being followed by ECRI-MGD investigators to develop, select, and validate individual IGDIs; (d) provide examples of possible IGDIs and overview findings from the initial research on these measures; and (e) describe our Institute's vision for the intended use of IGDIs in a comprehensive, data-based decision-making assessment model.

Purpose and Conceptual Foundation of Individual Growth and Development Indicators

In other work (ECRI-MGD Technical Report 1), we have described the relation between America's growing movement toward educational accountability and the assessment of develop mental status for preschool and early elementary-aged children and their families. In that report, we argue that the conceptual and logistical support for involvement of these populations – particularly for preschool children and children with disabilities and their families – is still quite weak, leading to uneven participation and low-quality information. At the same time, we argue that a functional, easy-to-administer, and valid system for assessing growth and development, and for supporting the modification or addition of intervention services when needed, will prove essential to any increase in observed levels of student performance under any of the current accountability systems. In short, single-point "high-stakes" assessments of child performance will not, in and of themselves, lead to improvements in children's levels of performance; such changes will occur only

when parents, teachers, administrators, and others have access to information, both over time and long before the high-stakes evaluation, that helps them improve the quality of services – and thus the outcomes – for all children participating in the educational system (c.f., Deno, 1985; Good & Kaminski, 1996).

It appears that a small set of common child and family indicators, based on robust and widely-held general outcomes, would assist the early intervention and early childhood special education community, would be consistent with calls for greater accountability in education and human services, and could respect and contribute to the unique perspectives of special education and related services for young children and their families. Such a system could:

- Maintain an individualized focus on current skills and levels of developmental functioning for children and families, and factors that promote growth and development;
- Place the growth and development of each child within a broader, more consistent psychometric context (or "common metric") of important, more widely-held values for child and family outcomes;
- Directly support the spirit and intent of Individualized Family Service Plans and Individualized Education Plans, providing families and practitioners information and measurement strategies that help operationalize many (but not all) long-term goals, and supporting creativity, individualization, and a focus on effectiveness of short-term objectives to reach these goals;
- As a natural by-product of individual assessment and program planning, gather information that can be "rolled up" or aggregated across individual children and families to provide useful information about groups of children in classrooms, programs, districts, or even States.

This vision is driven by a commitment to understanding and appreciating individual children's and families' growth and development over time, to accepting bottom-up implementation of innovations, and to an active and on-going problem-solving approach to designing and evaluating interventions for individual children and programs for groups of children. To achieve this goal, a comprehensive program performance measurement system must have six essential features, including:

1. *A small number of common measures for different age groups* of children and their families, with links in measures across ages, reflecting levels of development and developmental expectations, as well as the ways in which children of different ages and their families receive services;
2. *Adaptations of common measures for specific subgroups of children and families* that, because of personal characteristics (e.g., sensory impairments), demographic factors (e.g., poverty), or cultural variables (e.g., native language other than English), require unique information regarding development and growth over time;

3. *Specific links to systematic, well-developed intervention-referenced tools and procedures to help evaluate and plan changes in intervention*, including description of the supports, experiences, services, and treatments that a child and family currently receive, and guidelines or procedures for using this information for planning changes to further enhance the growth and development of that child and/or family;

4. *Clear guidelines (or decision rules) for monitoring growth and development*, helping decide when the rate of growth and development is less than desired, and for considering intervention-referenced information to plan changes for children and families;

5. *Focus on work samples, in vivo behaviors, and other "authentic" assessments* to ensure that measures of growth and development are meaningful and representative of functional skills in the child's life;

6. *Active and wide-reaching dissemination of information* about this system to parents, inservice practitioners and administrators, policy makers, researchers, and preservice trainers to initiate its wide-scale adoption, and ongoing technical assistance and training to support its implementation.

One of the key foundations to this comprehensive assessment system is the development of dynamic measures of individual child and family development. These measures must sample relevant domains of development across the desired age-range, must be sensitive to growth and change over relatively short periods of time, must be cheap and easy to administer and interpret, and must lead to (and support) evaluations of the effects of ongoing intervention efforts as well as planned variations.

Recently, Fuchs and Deno (1991) have distinguished "general outcome measures" from other types of repeated assessments intended to describe growth or monitor progress for individual children (particularly an approach they label "specific subskill mastery"). Simply put, general outcome measures are "reliable, valid, and efficient procedures for obtaining student performance data to evaluate their instructional programs....the two most salient features of measuring general outcome indicators are (a) the assessment of proficiency on the global outcomes toward which the entire curriculum is directed, and (b) the reliance on a standardized, prescriptive measurement methodology that produces critical indicators of performance" (Fuchs & Deno, 1991, pp. 488-493).

We believe that a general outcome measurement approach has much to offer the assessment of individual growth and development among young children and their families. First, a general outcome measure approach produces data that describes *rate of growth* across developmental or chronological periods, allowing parents and professionals to judge not only proximal intervention effectiveness, but also the extent to which intervention services are increasing the probability of desired long-term outcomes. Second, a general outcomes measurement approach increases the probability that professionals will use, and parents will have access to, psychometrically sound assessment results. By selecting a smaller set of specific measures, care and attention can be devoted design, evaluation, and improvement of each measure prior to application. Finally, a general outcomes measurement ap-

proach yields data that are comparable across time, children, and programs. These data are therefore more useful for efforts to "roll up" or aggregate individual progress data to describe group- or program-level outcomes.

While continuous progress measures are widely available in a variety of academic domains for elementary and secondary school-aged students (Shinn, 1992), and while some few instances of continuous progress measures of this type exist currently for young children and their families (Priest & McConnell, 1994), to date there are not data collection and assessment protocols, nor the foundational conceptual and empirical support, for comprehensive assessment of growth and development of children birth to age 8 and their families. Development of *individual growth and development indicators* under the auspices of this Research Institute is intended to begin efforts to provide the foundations and logistical supports for extending this type of assessment to young children and their families.

Criteria or Desirable Features for Individual Growth and Development Indicators

Development of individual growth and development indicators is guided by almost 20 years of research, development, application, and evaluation of similar measures for older children and other areas of student performance. Beginning in the mid-1970s, Stanley Deno and Phyllis Mirkin at the University of Minnesota began development of measures of continuous progress in academic domains (Deno & Mirkin, 1977). This work quickly led to initial development and evaluation of *curriculum-based measures* (Deno, 1985; Deno, 1986; Deno, Mirkin & Chaing, 1982), and to a rich and still-expanding literature regarding the development, evaluation, application, and interpretation of these measures, both alone and as part of instructional planning and program evaluation efforts (Fuchs & Fuchs, 1986; Fuchs, Fuchs & Hamlett, 1994; Good & Shinn, 1990; Marston & Magnusson, 1987; Shinn, Good, Knutson, Tilly & Collins, 1992; Shinn, Good & Stein, 1989; Walker et al., 1987). This body of research, and the expertise of individuals that have contributed to it, serves as an important conceptual and logistical foundation for our current efforts to develop individual growth and development indicators for children birth to age eight and their families.

Design standards for general outcome measures used for continuous progress monitoring were first articulated by Deno, Mirkin, and Chaing (1982). While specific criteria have shifted some and analytic procedures associated with these standards have matured, these standards remain fairly robust today. Specifically, individual growth and development indicators (like other general outcome measures, including curriculum-based measures) should be evaluated along six broad features or

characteristics, including the extent to which they: (a) measure key elements of important outcomes for children; (b) can be used efficiently and economically; (c) are standardized and replicable; (d) rely on generalized or "authentic" child behaviors; (e) are technically adequate; and (f) are sensitive to growth and change over time and to effects of intervention.

Measurement of Key Elements

General outcome measures (c.f., Fuchs & Deno, 1991) are, as stated, *general*: That is, rather than representing child performance on a specific task or skill, performance on these general outcome measures typically reflects child performance across a cluster of skills, from simple to complex. To produce a robust measure of general outcome, and a measure that is likely to show both correlations with other criterion measures and important social outcomes, assessment strategies must sample across various key elements of the developmental area of interest.

In applications with older children, researchers have demonstrated the reliability, validity, and utility of an oral reading measure (typically, words read correctly from text in one minute; see Deno, 1985). Child performance on this measure is assumed to include performance or mastery of elemental skills (e.g., letter-sound correspondence, left-to-right decoding, word attack skills) as well as more complex skills (e.g., comprehension and anticipation of content, reading for expression).

Similarly, initial efforts to identify general outcome measures in development[1] domains for preschool children have sampled across key elements. Priest (1998) gathered data of verbal communication, scoring children during free-play activities in classroom settings. This work produced a general measure – the proportion of 10-sec intervals where the child produced utterances of four or more words – that described levels of development. Other elements of performance on this measure would likely include articulation, vocabulary, semantics, syntactics, as well as social skills, play skills, motor skills, etc.

We assume that identification of key elements for any general outcome will assist in two ways. First, pragmatically, identification of these key elements assists in measure design and evaluation. As described on the next section, identification of these key elements is an important part of selecting data collection formats for sampling child behavior. Additionally, identification of key elements gives some direction to the selection of specific measures to describe child growth and development. Second, identification of key elements helps relate performance on growth and development indicators to the *Exploring Solutions Assessments* (see ECRI-MGD Technical Report 5), as well as efforts to describe and evaluate interventions.

It should be noted, however, that these key elements are not necessarily sampled singly in any individual growth and development indicator. As noted earlier, we are trying to create general outcome measures that are related to, but somewhat different from, critical skills mastery approaches that focus more specifically on these elemental features.

Efficiency and Economy

In application, individual growth and development indicators will be used to screen large numbers of children, as well as to monitor changes in development (through repeated assessment) over short and long periods of time. As a result, it is essential that these measures be quick to administer and score, that they produce data that are easy to manage and interpret, and that they require few consumable or unusual materials and little staff time. In development, when two measures produce data that are otherwise similar, the more easily and cheaply administered and used measure will be selected.

Standardized and Replicable

In application, individual growth and development indicators will be used across children, by multiple evaluators (including professionals and paraprofessionals), and across time. To describe growth and development, either in the individual case or for groups and programs, the data derived from these administrations must be comparable. Further, to ensure adequate levels of reliability and validity, variance due to examiner perspectives or administration procedures must be reduced. Thus, it is essential that all individual growth and development indicators be standardized with respect to administration procedures and scoring. In this way, replicability of these measures will be maximized.

Generalized or Authentic Child Behavior

For both empirical and face validity reasons, we assume that individual growth and development indicators should be drawn, to the maximum extent possible, from child behaviors that occur in (or are highly similar to behaviors that occur in) naturalistic settings. Empirically, naturalistic child behavior is the criterion "general outcome" to which these measures relate. Given conventional wisdom regarding the generalizability of behavior sampled in structured testing situations for younger children (e.g., Shepard, Kagan & Wurtz, 1998), we assume that child behavior sampled in more naturalistic settings will produce more robust and appropriate measures of overall development. Additionally, the fields of early childhood education specifically and education generally currently value those assessments that represent a child's "performance" in "authentic" tasks. Thus, to the extent that otherwise-adequate data can be derived in settings and ways that end-users judge desirable, we assume that implementation and application will be enhanced (McConnell, 1994).

Technical Adequacy

Individual growth and development indicators will be used to make decisions about individual children, including decisions about whether special education services are warranted, decisions about whether rates of growth and development are at desired rates, decisions about whether changes in intervention services are needed, and decisions about whether new or revised interventions are addressing specific needs. Thus, it is essential that the data sources supporting these decisions meet typical standards for reliability and validity of assessment data (APA, 1992; McClean, Bailey, & Wolery, 1997). Specific reliability standards for individual growth and development indicators will include inter-rater agreement, internal consistency (for measures with discrete items), temporal consistency, alternate forms reliability, and others. Validity standards will include discriminability by age and disability status, concurrent validity with criterion measures, predictive validity, and consequential validity with respect to *Exploring Solutions Assessments*.

Sensitive to Growth and Change Over Time

In addition to standard forms of psychometric validity, individual growth and development indicators must also demonstrate sensitivity to growth under two different conditions. First, to meet the basic test of general outcome measures for assessing development, individual growth and development indicators must be sensitive to changes over time (or development) for individual children. In particular, these measures must be sensitive to changes on the magnitude that occasion needs to begin formal or informal intervention services. At a minimum, then, we assume that individual growth and development indicators must be sensitive to typical developmental changes over months.

Additionally, individual growth and development indicators are intended to directly support the design and evaluation of individualized developmental or educational intervention. As such, these measures must be sensitive to changes due to intervention in the range that professionals and parents must evaluate intervention effectiveness. At a minimum, then, we assume that individual growth and development indicators must be sensitive to changes due to intervention over weeks.

Research Plans and Strategies for Formative Process of IGDI Development

Development of individual growth and development indicators also follows from empirical and logistical outcomes of earlier work developing curriculum-based measures (e.g., Deno, Mirkin & Chaing, 1982) and dynamic indicators of early literacy skills (Kaminski & Good, 1996). However, unlike earlier work in academic domains, the current activity focuses on child growth within *developmental* domains. While these domains are well-understood by teachers and researchers in early childhood education and early childhood special education, it appears that these domains are less well-defined and clear than "reading, writing, and arithmetic." Thus, the process of developing individual growth and development indicators has been adapted somewhat to ensure that our work is in areas, and develops measures, that are valued by likely end-users.

The overall logic model for developing individual growth and development indicators follows from general definition of areas or domains of concern – and social validation of the importance of these areas or domains – through a systematic research and development process intended to develop and evaluate measures that contain all of the desirable characteristics needed for this growth and development approach to assessment. The logic model has five steps, including (a) identifying general outcomes, (b) identifying key elements for each of these outcomes, (c) developing data collection formats for sampling child behavior related to each of these outcomes, (d) developing and pilot-testing scoring procedures to derive psychometrically sound growth and development data from these data collection efforts, and (e) establishing the technical adequacy and norms for these measures, including direct assessment of measures' sensitivity over time and contributions to treatment development.

Identify General Outcomes

ECRI-MGD Technical Report 2, Selection of General Growth Outcomes for Children between Birth and Age Eight, and ECRI-MGD Technical Report 3, National Survey to Validate General Growth Outcomes for Children between Birth and Age Eight, describe our efforts to identify and validate socially a comprehensive but parsimonious list of general developmental outcomes for young children and their families. While the list developed here is, both substantively and stylistically, different from other efforts to describe general outcomes for children and families in this age group, the intent of this effort is similar to that of earlier efforts: We want to describe those outcomes that parents, professionals, research, and community standards believe are *important* for young children, and that represent the desired results of informal or formal supports to developmental processes of the early childhood years.

These outcomes, and the survey data that offer evidence of their acceptability and appropriateness to a broad array of constituents, are central to our efforts to build a comprehensive assessment model of growth and development for young children and their families. These outcomes define the content, or universe, that more dynamic indicators must address.

To support empirical analyses of this match between what our community desires for young children and what the individual growth and development indicators measure over short and long periods of time, we must next operationalize each general outcome. Thus, as work on individual growth and development indicators associated with one particular general outcome begins, researchers must first identify one or more criterion measures for ongoing research. These criterion measures are selected, based on empirical review and expert opinion, to serve singly or in combination as the best and most direct estimates or measures for a particular general growth outcome.

Identify Key Elements For Each Outcome

Identification of key elements, as described earlier, assists both in measurement design and evaluation, as well as in selection of specific measures of child growth and development. In our work to date, key elements have been identified by reviewing empirical research, as well as existing assessment instruments and intervention curricula, to identify discrete components of development, or clusters of child behaviors, that are associated with and assumed to be essential to status in the more general domain. This is *not* akin to identifying specific subskills for monitoring mastery (c.f., Fuchs & Deno, 1991); rather, identification of key elements adds more detail and information to the definition of a general growth outcome for instrumentation and evaluation.[1]

Develop Data Collection Formats

Next in the development cycle is to identify situations or settings in which we can sample child performance related to the general outcome. While there is a close relation between this step and later development of specific scoring procedures, the emphasis here is on *how* to measure child performance in a general outcome area.

Several of the criteria for data collection formats relate specifically to the desirable features of IGDIs generally. First, data collection procedures must be *efficient and cost-effective*; procedures that employ available resources (including staff and materials) and make less intrusion on program operation are preferred. Second, data collection procedures should emphasize sampling child performance in naturalistic settings, producing data that end-users will judge to be *authentic*. Third, data collection formats should be standardized, both in terms of administrator behavior and with respect to situational variables that may influence child performance. The goal here is to reduce variance due to assessment, and to increase variance due to a child's current level of growth and development within a general outcome area.

Data collection formats must also have other features related to their intended use. These formats must occasion the desired types of responses from participating children. While not trying to reduce variance across children or time in this assessment, it is essential that situational and other features of the data collection format elicit the desired types of behavior or responses from participating children.

Second, these formats must be sufficiently engaging to participating children that the children will attempt or produce desired behaviors or responses. In many instances, formats will need to be tailored for different age groups or developmental levels; for instance, the types of activities that will be engaging to infants, and will elicit communicative behaviors, will be quite different from those activities that are engaging and elicit communicative behavior from early elementary students.

Last, these data collection formats must be repeatable, and thus available for frequent assessment. This means that either alternate forms or activities must be available, or that data are collected in activities that naturally repeat in the everyday lives of young children and families.

Develop and Pilot Measures

Related to, but separate from, efforts to sample child behaviors, we next will develop specific scoring procedures and measures to describe child growth and development. Again, these scoring procedures must contribute to development of IGDIs with all desirable features: These scores must be based on standardized procedures, yield reliable data (across scorers, forms, or situational variations), and be easy and cost-effective to produce (i.e., both time-efficient and easy to teach to scorers). Additionally, these scores must describe "growth and development" of young children and their families: They must discriminate between younger and older children developing similarly, they must correlate with criterion measures of the general outcome area, and they must be sensitive to changes within children over relatively short periods of time, either when intervention is provided or not.

Procedurally, pilot measures are selected based on logical analyses of general outcomes measurement, past research on general and/or cross-age measures of development, clinical judgment, and a focused series of hypothesis-generating and testing efforts (i.e., define a measure, apply the measure on a select number of cases, examine distributions and reliability coefficients, revise and/or expand test). The product of this effort is expected to be two to five specific measures that appear to be good candidates for individual growth and development indicators.

Establish technical adequacy of measures, including sensitivity and treatment validity

After isolating two to five possible measures, data are collected for a moderate-sized sample of children within a particular age range to begin estimation of formal psychometric properties of these measures. Research questions for this phase of work include those of reliability (e.g., inter-rater agreement, test-retest reliability, alternate forms reliability, and/or internal consistency), and validity (e.g., concurrent validity with criterion measures, known-groups discrimination for different groups [age, disability status], and predictive validity with other IGDI and criterion measures). In addition, formal evaluations of staff training, administration time and costs, and data management procedures will be conducted in this phase.

Last, and related to both development and evaluation of *Exploring-Solutions Assessments* (see ECRI-MGD Technical Report 5) and our overall monitoring and problem-solving model (see ECRI-MGD Technical Report 6), measures will be tested with likely-effective interventions. IGDIs will be judged to have adequate treatment validity if, and only if, they can be used to document the effectiveness of intervention services for young children with disabilities and their families.

Special Issues, Examples, and Results to Date, Individual Growth and Development Indicators for Three Age Groups

The ECRI-MGD workscope calls for development, evaluation, and field-testing of individual growth and development indicators for up to 15 general outcome statements (see ECRI-MGD Technical Report 2) in each of three overlapping age groups, including (a) infants and toddlers (birth to 36 months), preschoolers (30 months to 66 months), and (c) early elementary-aged children (60 months to the end of third grade). ECRI-MGD's research and development efforts for these indicators are further organized by site, with each participating university assuming primary responsibility for initial development and evaluation in an age group where they have special expertise: Researchers at the Juniper Gardens Children's Project (University of Kansas) site, led by Dr. Judith Carta, are developing infant IGDIs, a team led by Dr. Scott McConnell at the University of Minnesota is building preschool IGDIs, and Dr. Roland Good leads the team of investigators at the University of Oregon developing early elementary IGDIs.

These three research teams are following a common research and development plan, as described above, and are working in close collaboration to maximize consistency, whenever possible, across data collection formats and scoring procedures. To initiate this research and development process, and to facilitate development of common perspectives and procedures across age groups and teams, initial

efforts were directed toward developing and evaluating data collection and scoring procedures for a single common general outcome statement: *Child uses gestures, sounds, words, or sentences to convey wants and needs or to express meaning to others.* In this section, we describe special requirements and considerations for IGDI assessment in each of the three age groups, and present results of our initial efforts to develop and evaluate individual growth and development indicators for this one general outcome statement.

Special Requirements and Considerations for Assessing Growth and Development for Infants and Toddlers, Preschoolers, and Early Elementary Students

We expect significant variations in typical behavior, characteristics, and service delivery features across the chronological ages in our targeted group of children (i.e., birth to age 8). In some instances, these variations represent the very factors that we are trying to assess; for instance, we expect virtually all children to develop new skills and greater facility in many behaviors and areas of developmental competence across this eight-year span of time. However, many aspects of these variations represent other, extraneous variation ("noise" or error) in the assessment process that must be controlled or considered in designing any reliable and valid assessment. For instance, many children will develop from nonlinguistic to linguistic communication during this time. Additionally, children at different ages will spend time in different settings. Similarly, data collection formats that will be engaging (and thus appropriate) for children at one age or developmental level will be worthless for children of other ages or levels of development. Factors like these must be considered in the development of data collection and scoring formats for individual growth and development indicators that span this large period of chronological age. In this section, we describe some of the more salient requirements and considerations that must be considered for assessing growth and development in each of the three age groups.

INFANTS AND TODDLERS

Developmental considerations. Developing any type of valid assessments for very young children poses a special set of challenges. As we began the process of constructing indicators for measuring the growth and development of infants and toddlers in communication, we were aware of the following:

1. The notorious lack of predictive power of infant assessments based on single measurement occasions.
2. The realization that infants/toddlers have limited capacity to remain engaged in tasks for very long periods.
3. The fact that direct assessment of many infants/toddlers will often require the presence and participation of a parent or other familiar caregiver.

Indeed, any assessment of infants and toddlers, with or without disabilities, must balance the requirements and intent of formal assessment (e.g., standardization of stimuli and situational variables, standard assessment of child responses) with the characteristics and behavioral characteristics of this age group (e.g., quick behavioral state transitions, limited repertoires, and limited responses to verbal or situational instructions and structure). To some extent, we assume characteristics of assessment situations and measures will be similar across developmental domains for this age group, but that developmental considerations must be factored in to the development of all data collection and assessment protocols.

Service delivery considerations. Other considerations for test development for infants and toddlers relate to the types of settings in which these children spend most of their time. Unlike older children, very young children are not often found in settings with other same-aged children. The most typical place in which infants and toddlers are found is in their homes, in relatives' care or in home day care settings. Use of IGDIs in these settings might require parents, family members or other caregivers to make simple reports or observational assessments of their children. Accordingly, IGDI applications in these settings will require protocols that were very easy to use.

An alternative service delivery setting in which individual growth and development indicators might be used for regular screening of very young children might be through regular health care checks during visits to pediatricians or other health care professionals. Because time is at a premium during these visits, assessments for use in these settings must be very brief and easy to score.

If infants and toddlers receive care in group settings, such in child care or early education settings, they are typically run by private, not-for-profit organizations and not by public schools. The most typical staff person in these settings holds at most a bachelors' degree and has limited professional experience. As a result, experience in engaging children or assessing infants and toddlers is minimal. Hence, minimizing the complexity of test administration is required.

Sources of data. One final area in which developmental considerations had a bearing on the design of assessments was the source of the data. As noted earlier, very young children are not as likely as older children (all things considered) to produce specific behavior or responses to specific stimuli on verbal or situational demand. Additionally, while we assume that infant and toddler behavior may be more variable within and across individuals than that of older children, and while we assume that in many domains infant behavior is "emergent" rather than fully developed, we also assume that there are some consistencies in the behavior of these very young children. Taken together, we assume that infant and toddler behavior (and development) can best be assessed in situations, settings, or formats that capture natural variations, and levels of performance, as accurately as possible. Therefore, assessments for these youngest participants need to be designed in ways that children's behavior is either directly observed during naturalistic interactions or is obtained by reports of children's performance provided by their parents or familiar caregivers.

PRESCHOOLERS

Preschool IGDIs are intended for children between the chronological ages of 30 months and 6 years of age, with levels of developmental performance ranging from individual children with moderate to severe disabilities to children performing at or above the "normal" level. We have begun development of IGDIs for this age group with attention to broad measures of development, including typical behavior and settings of participation, for children this age.

Developmental considerations. We expect typical behavior to include increasing amounts of competence and complexity of responses in relatively unstructured, child-directed settings. For instance, children developing typically and those with mild to moderate delays in this age range will communicate with peers during free play, will initiate contact and interaction with toys and other materials, will put on and take off articles of clothing, and will engage in other behaviors with little or no formal structure or prompts to do so. Similarly, we expect performance at any moment in most naturalistic settings to be multi-dimensional, representing increasing skills and competence in multiple developmental "domains" (c.f., Guralnick, 1992). Finally, we expect that child behavior, particularly in naturalistic settings, will occur in an ecobehavioral surround (c.f., Carta, Greenwood & Robinson, 1992); complete understanding, assessment, and analysis of child competence will require a complex assessment of both the behavior of interest and environmental variables influencing, and influenced by, that behavior (Bijou & Baer, 1961).

Service delivery considerations. Similarly, formal service delivery options for children in the preschool age range often include congregate care or educational settings where substantial time is devoted to unstructured or semi-structured child-initiated or child-directed play (Ostrosky, Skellenger, Odom, McConnell & Peterson, 1994). We expect children receiving special education services to be assessed in early childhood special education and regular education preschool programs serving 8 to 24 children per classroom, in similarly-sized specialized or generic day care settings, in smaller home- or family-based day care settings, or in homes with parents or other adult caregivers. Across these settings, however, we expect a fairly large portion of available time to be allocated to activities that (a) provide some degree of child direction and initiation, and (b) present opportunities for a range of child behaviors and demonstration of multiple competencies.

Sources of data. Sources of data for preschool-aged children can vary widely. Given the behavior and service delivery settings typical for many children this age, naturalistic observation (Halle & Sindelar, 1982) can be used extensively. Similarly, observational assessment can be conducted in semi-structured settings, where some degree of control is exerted over materials, play partners, and themes or activities; these quasi-naturalistic settings may control for some variations in child performance, but still provide samples of child behavior that correspond closely with performance in more unstructured settings (Priest & McConnell, 1994).

Parent and teacher reports of child behavior can also be employed with children in this age range. These adult reports of child behaviors have the advantage of expanding the time-base or behavioral sample available for assessment.

Finally, preschool-aged children can be tested directly, although typically this assessment must be conducted in individual or very small group (e.g., 2-3 children) settings. Direct testing with standardized stimuli has been used extensively in broad developmental assessments (e.g., Newborg, Stork, Wnek, Guidibaldi & Svinicki, 1984) to describe child performance in a single domain or area (e.g., the Peabody Picture Vocabulary Test), or to describe child performance on a wide range of social and developmental tasks. Individualized or small-group assessment can be highly structured and standardized (like typically used for intellectual assessments), or can be quite unstructured and play-like (e.g., Linder, 1993).

EARLY ELEMENTARY STUDENTS

Early elementary IGDIs are intended for children between the ages of 5 and 8 years, or roughly from prior to kindergarten entry through the end of second grade. These IGDIs continue the developmental assessment paradigm to the upper limits of traditional "early childhood" programs, and link these measures to child performance in initial years of grade school enrollment. As such, these IGDIs are expected to complement, and relate to, both continuous-progress measures of late preschool and early elementary school (Kaminski & Good, 1996) and to curriculum-based and other measures of formal academic performance (Deno, 1985).

Developmental considerations. Under conditions of typical development, we expect children's behavior during this age period to (a) continue to develop and elaborate in all domains, as noted in child-directed settings, and (b) demonstrate increasingly competent performance in formalized, structured, adult-initiated tasks like those found in academic instructional settings. Like for preschool-aged children, we expect children with mild to moderate levels of developmental delay also to continue development and elaboration of skills and competencies in both child- and adult-directed activities. Also, we expect performance in most naturalistic, and many academic, activities to be multi-dimensional, representing increasing skills and competence in multiple developmental domains. Finally, we expect that child behavior, particularly in naturalistic settings, will occur in an ecobehavioral surround (c.f., Carta et al., 1992); complete understanding, assessment, and analysis of child competence will require a complex assessment of both the behavior of interest and environmental variables influencing, and influenced by, that behavior (Bijou & Baer, 1961).

Service delivery considerations. Formal services, both regular and special education, will be compulsory for virtually all children in this age group, and will be provided most typically in classroom programs. We expect children in this age group to be enrolled in classes ranging in size from 8 to 30 students. Additionally, children in this age group may spend considerable time in congregate-care settings (for after-school or other day care), as well as in family and community settings.

Sources of data. Sources of data for elementary-aged children can vary widely. Under many conditions, naturalistic observation can still be used in classroom, day care, playground, home, and community settings (Greenwood, 1992; Greenwood, Carta, Arreaga-Mayer & Rager, 1991; Greenwood,

Carta & Atwater, 1991; Greenwood, Carta, Kamps & Arreaga-Mayer, 1989). Similarly, observational assessment can be conducted in semi-structured settings, with some degree of control over materials, interactive partners, and themes or activities.

Parent and teacher reports of child behavior can still be used with children in this aged range. At some level, however, child behavior in this age range may occur outside of the purview of adult monitoring, beginning to mitigate the overall usefulness of this data source.

Perhaps most distinctly, early elementary students can be tested directly, either in individual or group formats. Direct testing with standardized formats has been used extensively in developmental and academic domains. This assessment can be highly structured and standardized, or can be embedded into ongoing curricular activities.

Application Example: Developing Individual Growth and Development Indicators for "Using Gestures, Words, or Sentences to Communicate Wants and Needs or Express Meaning to Others"

As noted earlier, to facilitate consistency across the age groups and to develop a stronger shared sense of standards and procedures for proceeding with later R&D efforts, three age-based IGDI research and development teams have initiated work on the same general outcome statement. Given parents' and teachers' relative interest in communication outcomes (see ECRI-MGD Technical Report 3), our assumption that individual growth and development indicators in this general area would prove important to other end-users and evaluators, and the relatively high level of empirical research and assessment experience within the broad domain of communication and language, we elected to start with the following general outcome statement: *Child uses gestures, sounds, words, or sentences to convey wants and needs or to express meaning to others.* In this section, we will describe progress to date by the three age-group teams in five aspects of work related to developing and evaluating IGDIs for this outcome: (a) analysis of the key elements for this general outcome at a particular age level, (b) criterion measure selection, (c) data collection formats, (d) scoring procedures, and (e) initial results, including distributions of scores by age and correlations with criterion measures.

INFANTS

Key elements analysis. The first step in the IGDI development process was the identification of key elements (those behaviors related to outcomes) that represented essential sampling requirements to produce general outcome measures (Fuchs & Deno, 1991) within brief, periodic assessment occasions to measure rates of growth. In order to identify these elements or behaviors, we conducted a review of studies that described early expressive communication development; this review was drawn from our earlier, more comprehensive review of developmental outcomes and general outcome statements for infants and young children (see ECRI-MGD Technical Report 3). We also reviewed available assessments related to general communication domains. We specifically sought behaviors

that demonstrated growth over time, were amenable to intervention, could be measured reliably, and were related to the outcome of interest. The behaviors that best fit those criteria for children between 6 months and 36 months were the following:

1. social attention and coordinated attention,
2. gestures,
3. vocalizations
4. discrete words,
5. word combinations.

Data collection formats. Next, we were interested in developing and evaluating informally the capacity of different data collection formats for sampling child behaviors related to this general outcome, as well as the key elements described above. The purpose of this data collection format was to standardize administration across time and examiners, and to identify situations in which infants and toddlers would produce reasonable samples of child behavior for more detailed assessment and IGDI development. For infants and toddlers and this particular general outcome statement, three types of data collection formats have been developed to address the expressive communication IGDI: Communication Evoking Situations, Naturalistic Observation Situations, and Caregiver Communication Measure.

Communication evoking situations (CES) are semi-structured play situations designed to evoke various communication functions (i.e., "want more," "want help," and establishing joint attention). The basic format for this CES format was adapted from the *Communication Symbolic Behavior Scale* (Wetherby & Prizant, 1993) and the *Abridged Early Social Communication Scales* (Mundy, Hogan, & Doehring, 1996). The following factors were considered in developing these situations:

- They would evoke a broad distribution of the key elements of communication for children in the 6-36 months age range;
- They would interest children in the entire age range of 6 to 36 months;
- They would engage children quickly;
- They could be used repeatedly over time and still maintain children's interest;
- They would be safe and developmentally appropriate for all children in the age range.

A number of questions were posed about the design of the CES. Among these were: how much prompting should be employed to evoke communication behaviors? Should verbal and physical prompts be standardized? Who should prompt child to communicate (a parent or familiar caregiver, an impartial examiner, or a combination)? Should the same set of toys be used for the entire age range? Should the time frame for child responses be standardized or be free to vary? How many situations should be employed?

A pilot study was conducted to answer these questions and to narrow down the set of potential CES items. After examining a wide range of situations, a set of nine CES items was selected based on criteria described above. In addition, CES items were chosen that produced the broadest distribution of scores for children tested. A protocol was finalized for the administration and scoring of the nine CES situations and appears in Appendix A. These sessions are videotaped and scored from the tapes. We are using videotaping only for research version of the instrument and, in keeping with our desire to make these instruments are easy to use as possible, plan to move the assessment into a format that can be scored *in vivo*.

The *Naturalistic Observation Situations* (NOS) is a semi-structured play session in which a parent or familiar caregiver is asked to play with the child for 10 minutes with a toy selected to provoke a wide range of communication and play behaviors. Two alternate toys were selected – the Fisher-Price® House and Fisher-Price® Barn – for alternate forms of the assessment. The caregiver is given one of these toys and is asked to play with the child as s/he normally would. For the present time, the entire session is videotaped and behavioral measures are scored from the tape using the same scoring procedures used for the CES items. The NOS protocol appears in Appendix A.

The third data collection format is the *Caregiver Communication Measure* (CCM). This is a questionnaire of 100 items in which the parent or other familiar caregiver is queried about the child's communication. Items have been adapted from the *MacArthur Communication Development Inventory* (Fenson et al., 1993); current research by Hart and Risley (in press); the *Assessment, Evaluation, and Programming System* (Bricker, 1993); *CSBS Screening and Evaluation Tool: CSBS Developmental Profile* (Wetherby & Prizant, 1998), the *Denver Developmental Test-II* (Frankenburg, 1990); and the *Preschool Language Scale-3 Family Information Form* (Zimmerman et al., 1992). Most items are formatted as Yes/No responses to questions about the child's communication such as "Does your child make sounds such as gurgling or babbling when alone or when playing with toys?" Other items require the caregiver to indicate which of the gestures or words the child has been known to use, respond to, or say. For example: "Does your child talk about things that happened in the past such as: "My cup fall down." "I fell down." or "I falled down." Two alternate forms have been developed each with 100 items. They appear in Appendix A.

Scoring procedures. Key elements within the CES and NOS videotaped sessions are currently being scored using two types of procedures. The first system being used for research purposes employs an event-based observation system. The events recorded are the key elements of communication (see Definitions in Appendix D.) This system uses a "code sequence" format. Using this recording procedure, code letters that represent specific behaviors (i.e., units of child communication) are written in sequence as they occur (see codes at the end of definitions. Behavior combinations are noted by entering both codes (e.g., GV) and putting a circle around them. Advantages of this system are: (1) it permits the collection of information on sequence of behaviors that might be useful during instrument development; (2) its format can be translated into a computerized system that would allow for more efficiency in *in vivo* assessments.

Anticipating the instrument's use in the field, we are also testing a data recording procedure that utilizes a checklist format. In this system, data are recorded by checking off the appropriate category. We are working on strategies for recording behavior combinations (e.g., gesture + vocalization). Advantages to this method are that it facilitates data summary and may be easier for teachers and other practitioners to learn. The Definition of Communicative Categories, along with coding forms, appear in Appendix A.

Criterion measure selection. One final step in the development of IGDIs was selection of a criterion measure against which we would validate each IGDI format. The most important criterion for its selection was its "goodness of fit" with the general communication outcome we are addressing: "child can use gestures, words, or sentences to communicate wants and needs or express meaning to others." Other important selection criteria included: strength of psychometric properties (reliability, validity, internal consistency); diversity of standardization sample; extent of age range coverage to match infant/toddler IGDIs; availability of scores that could be used in criterion validity studies with IGDIs. The following measures were considered:

- Communication and Symbolic Behavior Scales (Wetherby & Prizant, 1993)
- MacArthur Communicative Development Inventories (Fenson et al., 1993)
- Reynell Developmental Language Scales (Reynell & Gruber, 1990)
- Preschool Language Scale-3 (Zimmerman, Steiner, & Pond, 1992)
- Sequenced Inventory of Communication Development (Hedrick, Prather, & Tobin, 1984)
- Receptive-Expressive Emergent Language Scale (Bzoch & League, 1991)

The Preschool Language Scale was chosen for a number of reasons, including psychometric properties of the instrument, its inclusion of children from low-income families in the standardization sample, the availability of adaptations for children with severe delays and physical or hearing impairments, the breadth of ages covered, and the fact that the PLS is being used as a criterion measure to validate the preschool IGDI.

Initial results. An initial study to examine the psychometric properties of the three IGDI data collection formats is currently underway. A ethnically and socioeconomically diverse sample of 30 children aged 5 months to 36 months sites have been recruited from child care sites in Kansas City and Lawrence, Kansas. Data are being collected on alternate forms of the Caregiver Communication Measure, the Naturalistic Observation Situation and the Communication Evoking Situations. The primary research questions being addressed in this study are the following:

1. Do the three data collection formats (CES, NOS, and Caregiver Communication Measure) produce the desired distribution of scores across the age span using the various scoring approaches? Do some methods of scoring yield better distributions than others

2. Can independent observers obtain adequate levels of interscorer agreement when scoring the CES and NOS protocols across the age span?
3. Are correlations of sufficient magnitude obtained on alternate forms of the measures?
4. Are correlations of sufficient magnitude obtained on each IGDI measure with the criterion measure (PLS)?

Data answering these questions will be used to select the strongest IGDI measures. This subset of measures will be used in an intensive study this summer in which children are measured repeatedly to examine their pattern of growth on IGDIs as well as the IGDI measures predictive validity with later criterion measures.

PRESCHOOLERS

Key elements analysis. Using information gathered for our initial development of general outcome statements (see ECRI-MGD Technical Report 3), the preschool IGDI team had a large number of items from empirical research, assessment instruments, and curriculum programs to describe child performance related to this broad general outcome statement. Based on review of these items, and confirmed by available reviews of typical language development for children in the 3 to 6 year age-range, we identified four broad classes of key elements:

1. Production of discrete words (vocabulary),
2. Production of word combinations or sentences,
3. Fluency in grammar and mechanics, and
4. Recounting events or experiences, and telling stories (related to themes and information)

These key elements build on those identified for infants and toddlers, including two key elements that are identical (i.e., production of discrete words and production of word combinations) that are expected to continue development and elaboration during the targeted age range, and two elements that are expected to emerge and continue development in typical patterns observed in this age range.

Data collection formats. Given typical service delivery settings for this age children, and based on earlier work conducted by our research group (Madyun, 1996; Priest, 1998; Priest & McConnell, 1996) and other colleagues associated with ECRI-MGD (Kaminski & Good, 1996), we identified three data collection formats for sampling preschool children related to this general outcome: Picture Naming, Story-Telling, and Semi-Structured Play.

The Picture Naming task is adapted from Kaminski and Good's (1996) earlier work on dynamic indicators of early literacy skills. In this task, children are tested individually by an adult examiner. For preschoolers, we have prepared three series of 30 color photographs and line drawings (based on pilot test comparisons of these formats with black-and-white line drawings) of common objects,

including those found in homes and classrooms (e.g., fork, clock), body parts (e.g., ear, knee), and objects in the environment (e.g., tree, ladder). The examiner briefly describes and demonstrates the task for the child, and then presents pictures one at a time, recording correct or incorrect responses and presenting a new stimulus after every child response or 5 secs of nonresponse. Detailed description of the protocol for this data collection format is presented in Appendix B.

Story-telling is the second data collection format developed for preschool children. This format is also administered individually to a child by an adult examiner; again, the child and adult are seated at a child-sized table. Materials for this format include thematic, sociodramatic figures and toys commonly found in preschool classrooms (e.g., Fisher-Price® Fire House or Play-Skool® Barn). The examiner labels materials available for the child, and tells a very brief illustrative story. The child is then instructed to tell a story, using the materials present, that is "longer and better" than the examiner's. The examiner prompts the child to continue the story, or tell that it is completed, after 10 secs of nonresponse or two full minutes. Three story-telling opportunities are provided per occasion (see Appendix B for complete administration directions).

Pilot testing of this story-telling task indicated that children developing typically across the age range engaged the task, followed the examiner's directions, and produced some level of response. These initial evaluations also suggested, however, considerable variation in examiner behavior (particularly delivery of prompts for more child response, and request for elaboration of specific story elements), producing unwanted variations in observed child behavior. Based on this field-test, we further standardized examiner reactions and prompts, and introduced repeated delivery of new story themes to ensure a reasonable estimate of child behavior for further assessment.

Third, based on work by Priest (1998), we developed a semi-structured play setting for sampling child behavior. In this data collection format, two same-sex and approximately same-aged children are placed in a small, well-defined area that has commonly available toys or other materials known to elicit social communicative behavior (e.g., dolls, Legos®). Materials are described to the children, and several general play themes are presented. The examiner then asks the children to play together for up to 10 minutes. The examiner maintains children in the semi-structured play setting and monitors child behaviors, but otherwise delivers no prompts or feedback.

Priest's (1998) earlier work sampled child behavior during classroom freeplay activities, where children typically had a range of activity, material, and play partner options. Priest's analyses indicated within- and between-session variability; anecdotally, this appeared to be due in large measure to variations in the environmental variables. Based on this analyses, we pilot-tested several different formats for controlling play setting variables (including materials and access to a partner) for this assessment purpose, arriving at the current data collection format. Administration procedures are presented in Appendix B.

Criterion measure selection. Using criteria similar to those described for the infant IGDI, as well as those described by Deno, Mirkin, and Chaing (1982), we sought criterion measures that were likely to be judged as measures related to the general outcome statement, and that had strong psychometric properties, a broad and diverse standardization sample, an appropriate age-range, and derived scores that could be used in parametric analyses with IGDI measures. Based on our review, and discussion with special education researchers with expertise in the area of language and communicative development, we selected two measures: The *Preschool Language Scale* (Zimmerman et. al, 1992) and the *Peabody Picture Vocabulary Test, 3d Edition* (PPVT-3, Dunn & Dunn, 1997).

Scoring procedures At the time this report was prepared (April, 1998), we were engaged actively in developing and evaluating different scoring procedures for each of the three data collection formats, particularly for story-telling and semi-structured play samples that produced more naturalistic child behaviors. Child participants to data have had their performance videotaped, and all scores are being derived from these permanent records. Analyses of these scoring procedures, described in the following section, were intended to identify scores that (a) demonstrated good reliability, (b) produced variation across ages and developmental levels of participating children, and (c) correlated with criterion measures of child performance.

For the Picture-Naming task, scores to date include the total number of pictures identified in a one-minute sample. No other estimates are likely to be derived from this data collection format.

For Story-Telling, we are evaluating measures of the story specifically, and measures of child verbal production more generally. Related specifically to the story, we are measuring (a) length of story, in seconds and in words; (b) number of unique nouns; (c) number of phrases and sentences; and (d) number of objects (in play set) named. Additionally, we are adapting measures of language production from other efforts, including the event-based observation system employed with infants (see Appendix B), and Priest's (1998) direct observational assessment of 10-second intervals in which the child produces utterances of three or fewer and four or more words (see Appendix B). For the Semi-Structured Play format, we are using Priest's (1998) observational protocol, as well as the event-based observation system employed with infants.

Initial results. In addition to a series of preliminary studies (Kaminski & Good, 1996; Madyun, 1996; Priest, 1998; Priest & McConnell, 1994), and a series of pilot studies of data collection formats for use with preschool children, we have recently completed data collection for an initial sample to develop initial IGDI administration and scoring procedures for this "expressing meaning" outcome for preschool-aged children. The sample is ethnically and economically diverse (but drawn from a university-based day care center), and includes 41 children from 30 to 64 months of age. All participants completed both the PPVT-3 and PLS, as well as Picture Naming, Story-Telling, and Semi-Structured Play. All administrations of Story-Telling and Semi-Structured Play sessions were videotaped, typically using a wireless microphone recorded to an adjacent video camcorder, for later scoring. Primary research questions for this study include:

1. To what extent do identified data collection procedures yield scores that increase as a function of chronological age for children 30 to 60 months of age?
2. To what extent do identified data collection procedures correlate with one or more criterion measures?
3. For observational measures, to what extent do two independent observers produce similar scores when observing the same sample of behavior?
4. To what extent do identified data collection procedures yield scores that increase over time within subjects?

Analyses from the current sample will be used primarily to address the first three questions. While more detailed analyses will be presented to reviewers in mid-May 1998, the earliest analyses of these data provide some preliminary information. These analyses are based on the first 11 participants, 40 to 60 months of age, completing all IGDI data collection formats and both criterion measures (scores are presented only for Picture Naming and Semi-Structured Play tasks here). Table 1 presents child-level performance on all measures. In general, scores for Picture Naming and Semi-Structured Play show variations across participants, with no evidence of individuals obtaining absolute floor- or ceiling-level scores on any measure.

Table 2 presents product-moment correlations among select measures from this initial sample. Due to differences in chronological age that crossed norm-groups for the PPVT, raw scores for this measure were used. Correlations between age or criterion measures (i.e., PPVT and PLS) and three early IGDI measures demonstrated a range of correlations at low to moderate levels of magnitude, with higher and significant correlations between both criterion measures and performance on the Picture Naming task. Additionally, while lower and not significant for this sample, correlations with measures from the semi-structured play group (particularly *four or more word utterances*) were promising. These analyses will be expanded substantially, both for a larger and more age-diverse sample and across additional measures, in Spring 1998.

Table 1: Means and Standard Deviations of Prospective IGDIs and Criterion Measures

Child ID	Age Months	Sex	PPVT (raw)	PLS (standard)	PicNam	≤3 Words	≥4 Words
M01016	40	F	90	150	21	6	25
M01018	41	M	41	76	12	21	16
M01023	42	F	70	115	18	7	48
M01013	49	F	64	100	22	5	15
M01011	49	M	58	97	19	20	41
M01001	52	F	53	93	18	19	18
M01008	52	F	75	134	22	15	43
M01012	54	F	83	119	24	18	31
M01004	57	F	82	122	19	14	21
M01005	59	F	92	123	26	15	19
M01002	60	F	52	104	18	17	8
Total Group							
Mean	—	—	69.09	112.09	19.91	14.27	25.91
SD	—	—	16.87	20.73	3.73	5.75	13.06

Note. PPVT = Peabody Picture Vocabulary Test—Third Ed.; PLS = Preschool Language Scale - 3; PicNam = Picture naming situation; < 3 = # intervals in which child produced an utterance of 3 or fewer words; > 4 = # intervals in which child produced an utterance of 4 or more words.

Table 2 Concurrent Criterion-Related Validity, Preschool IGDIs (N = 11)

	PPVT (raw)	PLS (standard)	Picture Naming	≤3 Words	≥4 Words
Age	.18	.05	.46	.35	-.35
PPVT	—	.88**	.80**	-.45	.25
PLS		—	.64*	-.52	.32
Picture Naming			—	-.31	.14
≤3 Words				—	-.12

* $p < .05$. ** $p < .01$.

EARLY ELEMENTARY STUDENTS

Key elements analysis. Using information gathered in the development of general outcomes, as well as existing research on early literacy skills assessment (Kaminski & Good, 1996), as well as the substantial body of research on relations between language and communication competence and early school performance, the early elementary IGDI team identified four key elements of the "expressing meaning" general outcome; as we saw earlier, there is substantial overlap with analyses for the preschool-age group, reflecting the ongoing aggregation and elaboration of development. Specific key elements for early elementary students include:

1. Vocabulary, including production of discrete words
2. Extended discourse, including production of multi-word utterances, sentences, and longer thematic or expository statements
3. Narrative production tasks, such as telling stories, recounting events, or describing actions
4. Formal definitions of words, concepts, and actions, representing general store of knowledge.

Data collection formats. Given typical service delivery settings and developmental considerations for early elementary-aged children, we identified three individually-administered data collection formats for these children. Formats include picture naming fluency, narrative discourse through story retell, and picture description.

The Picture Naming Fluency task is similar to versions for preschool-aged children, as well as earlier work on dynamic indicators of early literacy development (Kaminski & Good, 1996). In this format, the child is presented line drawings of common nouns arranged in random order and printed six per line on 8.5 by 11 inch sheets of paper. Twenty alternate forms of this format have been developed, from a 1000-word picture pool. These pictures are drawn from preprimer to level 5 phonics workbooks of Scribner Reading series, pictures of common nouns from preprimer to level 5 of Scribner Reading series, pictures of words commonly used by 5-, 6-, and 7-year olds (Wepman & Hass, 1973), and pictures of words listed in the first three levels of the Harris-Jacobson Word List (Harris & Jocobson, 1972). (See Appendix H for administration instructions and samples.)

The Story Retell task is also completed individually. After brief instructions, the examiner reads a pre-primer story (one of 24 "Tom and Pippo" stories, from LadyBug Magazine, used with the publisher's permission) to the child. Each story consists of five single panels with a line-drawing and a one or two-sentence story element; these frames are placed on the table in front of the child. After completing the story, the examiner asks the child to look at the pictures in sequence and retell the story. Sample administration materials are presented in Appendix I.

The Picture Description story is similar to the story telling task for preschool-aged children. The child is presented a colorful, thematic picture and is told "I want you to tell me a story. I am going to show you a picture first, and then I want you to tell me a short story about what is happening in the picture. You will have 15 seconds to think about the story you will tell, and then have 1 minute to tell it. Try to tell a complete story and tell me when you are finished. Are there any questions?" Sample pictures and complete administration instructions are presented in Appendix J.

Criterion measure selection. After review of available tests for this age-group, and consideration of factors described in criterion-measure selection for the two younger age groups involved in this research, investigators selected three measures as criteria for early elementary-aged children: the Peabody Picture Vocabulary Test (PPVT-3), the Expressive Vocabulary Test (EVT; Williams, 1997), and the Test of Oral Language Development-Primary 3 (TOLD-3; Newcomer & Hammil, 1997).

Scoring procedures. Various scoring procedures are also under active development and evaluation for this age-group, with additional information available in May 1998. As with younger age groups, participants' responses to the two open-ended data collection formats (Story Retell and Picture Description) are audiotaped for later scoring.

For Picture Naming Fluency, scoring to data include the total number of pictures identified in a one-minute sample. For Story Retell, measures included total time (i.e., number of seconds retelling story from start to finish), number of word verbatim repeated from the original story (expressed as rate per minute and frequency), the total number of word said (total frequency and rate per minute), and the total number of functional language units (or ideas from story, expressed as total frequency). For Picture Description, scores to date include total words spoken over a full minute and per 10-sec interval; initial efforts to assess number of story grammar units produced little usable information, but future efforts may examine more specific story elements (e.g., t-units, parts of speech).

Initial results. As part of an ongoing effort to develop elementary IGDIs, the early elementary team is currently pilot-testing data collection formats and scoring procedures for a sample of 34 kindergarten through second-grade students in one elementary school. Children were tested on all data collection formats, as well as criterion measures, with IGDI measures repeated on three occasions. Data from this effort are currently being analyzed, and will be presented during the external site review in May 1998.

Table 3 Initial Results, Picture Naming Fluency and Criterion Measures, Kindergarten and First Grade

	McCarthy Scales	*Metropolitan Readiness Test*	*Stanford Diagnostic Reading Test*	*Rhode Island Pupil Identification Scale*	*Teacher Rating Scale*
Kindergarten	.67**	.74**	—	.52**	.75**
First Grade	.31		.42*	.15	.32*

In earlier work, the Picture Naming Fluency measure was administered to a sample of kindergarten to first-grade students; rather than current criterion measures, this study included a range of academic readiness and achievement tests commonly used with early elementary students. Results indicated that Picture Naming Fluency is correlated, at moderate to higher levels, with criterion measures administered to these students. Correlation coefficients are somewhat lower, with fewer significant relations, for first-grade students. Further analyses, including assessment of child performance over time, are expected to expand these findings. Additionally, reliability and validity data for the Story Retell and Picture Description tasks will be available during May 1998.

Using Individual Growth and Development Indicators in a Problem-Solving Assessment Model

As noted in the initial sections of this technical report, collection of individual growth and development indicators is one part of the comprehensive assessment and problem-solving model that is under development by investigators of the Early Childhood Research Institute on Measuring Growth and Development. As described more fully in ECRI-MGD Technical Report 6, this comprehensive model includes (a) ongoing assessment and monitoring of growth and development for individual children, either across developmental outcome areas or in areas of particular interest or concern; (b) application of formal decision rules to evaluate individual levels of development or rates of growth, as assessed by IGDIs; (c) when levels of development or rates of growth are less than desired, formal assessment of intervention status and generation of different new intervention revisions or options through formal *Exploring-Solutions Assessments* (see ECRI-MGD Technical Report 5); (d) implementation of new or revised interventions, when needed, with monitoring of fidelity or completeness of intended services and supports; and (e) ongoing monitoring and evaluation of intervention effectiveness with IGDIs that initially led to decisions to implement new or different interventions or supports. Thus, this comprehensive model is *formal*, in that it produces data and information of known features and psychometric characteristics for making decisions; *systematic* in that it prompts discrete decisions with clear, available standards or guidelines; *idiographic* because it focuses on the development and growth of individual children and families at specific points in time, and generates intervention or support recommendations (when needed) that address the specific wishes, preferences, strengths, and needs present for that individual and family; and *recursive* in that ongoing monitoring of growth and development allows parents and professionals to judge status at any point in time, and to evaluate continuously the efficacy of their plans and efforts.

As work progresses on the development and evaluation of specific IGDI measures, and as we begin to collect larger datasets of measures for individuals over extended periods of time, formal research and development efforts will turn to generation of *decision rules* for assisting parents and professionals in evaluating individual growth and development. Models for development, evaluation, and application of decision rules are available in the special education literature, including early and seminal work by Haring, Liberty, and White (1980), as well as more recent work on data from curriculum-based measurement (Deno, 1985; Fuchs et al., 1994). We expect these decision rules to provide for normative comparisons of individual child performance at one point or over extended periods of time, as well as intra-individual evaluations of rate of growth and progress toward desired long-term outcomes, using both relatively simple graphical analyses and (when needed) more detailed statistical analyses (Good & Shinn, 1990; Shinn, Good, & Stein, 1989).

We believe that this comprehensive assessment model will further support and expand family involvement, participation, and influence over the formal and informal services and supports that are provided to young children with disabilities. We expect the information generated by IGDIs to be direct, clear, and easy to interpret; we expect decision rules to be similarly clear and explicit; and we expect any possible alternatives generated by Exploring-Solutions Assessments to be strongly based in understandable information and data. With this information, we believe that families will have better and more timely information to monitor the growth and development of their children, as well as better information to guide (and evaluate) the work of professionals responsible for providing service. Conventional wisdom suggests that "information is power," and the information provided by this system will provide a strong and explicit basis for exercising the power and authority that families hold.

Finally, information generated by this comprehensive assessment system will contribute directly to ongoing efforts to develop, implement, and evaluate accountability systems for educational and other human service programs. IGDI data, consistent across children of similar ages and/or developmental levels, are directly available for aggregate-group analyses. Along with data from Exploring-Solutions Assessments, communities will have access to clearer information about the general and specific nature of services and supports provided to young children and their families, and the consequences of these services and supports for their growth and development.

Conclusion

The purpose of this paper is to provide a broad overview of the rationale for, and characteristics of, individual growth and development indicators for children birth to age eight and their families. To date, work on development of these individual growth and development indicators has included development of conceptual and theoretical foundations, articulation of general outcomes of interest, and initiation of formal research and development activities for a first set of indicators. Progress to date suggests that goals for ECRI-MGD, particularly those related to development of measures of individual growth and development, are attainable, and that it is reasonable to expect a comprehensive but parsimonious set of measures for children birth to age eight and their families. Without question, much research, development, field-testing, and dissemination remains; it is quite likely that the workscope for future years must be carefully focused and maintained to generate data and exemplars that are compelling to the field of early childhood education. With this focus, we expect to produce a set of immediately useful measures and procedures, as well as the foundation for further development, elaboration, and application of this comprehensive assessment system for young children and their families.

References

Bijou, S. W., & Baer, D. M. (1961). Child Development I: A systematic and empirical theory. New York: Appleton-Century-Croft.

Bricker, D. (1993). Assessment, Evaluation, and Programming System (AEPS) for infants and young children: Vol. 1. Measurement for birth to three years. Baltimore: Paul H. Brookes Publishing Co.

Bzoch, K. R., & League, R. (1991, 1971). Receptive Expressive Emergent Language Scale. Austin, TX: Pro: Ed.

Carta, J. J., Greenwood, C. R., & Robinson, S. L. Application of an eco-behavioral approach to the evaluation of early intervention programs, (pp. 1-68).

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52(3), 219-232.

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school pschologists. School Psychology Review, 15(3), 358-374.

Deno, S. L., Mirkin, P. K., & Chaing, B. (1982). Identifying vaild measures of reading. Exceptional Children, 49(1), 36-45.

Dunn, L. M. & Dunn, L. M. (1997). Peabody Picture Vocabulary Test (3d ed.). Circle Pines MN: AGS.

Fenson, L., Dale, P.S., Reznick, J. S., Thall, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. (1993). MacArthur Communicative Developmental Inventory (CDI). San Diego, CA: Singular Publishing Group, Inc.

Frankenburg, W.K., Dodds, J., Archer, P., Bresnick, B., Mashka, P., Edelman, N., & Shapiro, H. (1990). Denver Developmental Screening Test (Denver-II). Denver, CO: Denver Developmental Materials, Inc.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. Exceptional Children, 57, 488-500.

Fuchs, L. S., & Fuchs, D. (1986). Linking assessment to instructional intervention: An overview. School Psychology Review, 15(3), 318-323.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1994). Strengthening the connection between assessment and instructional planning with expert systems. Exceptional Children, 61(2), 138-146.

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/ prevention model of decision-making for early literacy skills. School Psychology Quarterly, 11(4), 1-11.

Good, R. H., & Shinn, M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. Behavioral Assessment, 12, 179-193.

Greenwood, C. R. (1992). Conceptual, methodological, and technological advances in classroom observational assessment, (pp. 1-37).

Greenwood, C. R., Carta, J. J., Arreaga-mayer, C., & Rager, A. (1991). The behavior analyst consulting model: Identifying and validating naturally effective instructional models. Journal of Behavioral Education, 1(2), 165-191.

Greenwood, C. R., Carta, J. J., & Atwater, J. (1991). Ecobehavioral analysis in the classroom: Review and implications. Journal of Behavioral Education, 1(1), 59-77.

Greenwood, C. R., Carta, J. J., Kamps, D., & Arreaga-Mayer, C. (1989). Ecobehavioral analysis of classroom instruction.

Guralnick, M. J. (1992). A hierarchical model for understanding children's peer-related social competence. In S. L. Odom, S. R. McConnell, & M. A. McEvoy (Eds.), Social competence of young children with disabilities: Issues and strategies for intervention, . Baltimore MD: Paul Brookes.

Halle, J. W., & Sindelar, P. T. (1982). Behavioral observation methodologies for early childhood education. Topics in Early Childhood Special Education, 2(1), 43-54.

Hart, B. & Risley, T. R. (in press). Learning to talk in America. Baltimore, MD: Brookes Publishing Co.

Hedrick, D. L., Prather, E., M., & Tobin, A. R. (1984). Seattle, WA: University of Washington Press.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. School Psychology Review, 25, 215-227.

Madyun, N. (1996). Preschool children's story-telling as a continuous progress measure of language development. Unpublished manuscript, Undergraduate Research Opportunity Program, University of Minnesota.

Marston, D., & Magnusson, D. (1987). Curriculum-Based assessment: District level implementation, Unpublished manuscript, Department of Special Education Minneapolis Public Schools.

McConnell, S. R. (1994). Social context, social validity, and program outcome in early intervention. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Herron, W. L. Heward, J. W. Eshleman, & T. A. Grossi (Eds.), Behavior analysis in education: Focus on measurably superior instruction, (pp. 75-85). Baltimore: Brooks/Cole.

Mundy, P., Hogan, A., & Doehring, P. (1996). A preliminary manual for the abridged Early Social Communication Scales (ESCS). [On-line]. Available: http://yin.psy.miami.edu:80/Child/Pmundy/manual.html.

Newborg, J., Stork, J. R., Wnek, L., Guidibaldi, J., & Svinicki, J. (1984). Battelle Developmental Inventory. Allen TX: DLM/Teaching Resources.

Ostrosky, M. M., Skellenger, A. C., Odom, S. L., McConnell, S. R., & Peterson, C. (1994). Teachers' schedules and actual time spent in activitites in preschool special education classes. Journal of Early Intervention, 18(1), 25-33.

Priest, J., & McConnell, S. (1994,). Continuous progress measures in early childhood special education: Initial development efforts. Paper presented at the Division of Early Childhood, Council for Exceptional Children.

Priest, J. S. (1998). Continuous progress measurement: Developing an observational instrument for preschoolers in early childhood special education classrooms. Unpublished Doctoral dissertation, University of Minnesota, Minneapolis.

Reuter, J. & Bickett, L. (1990). Kent Infant Development Scale, Second Ed. (KIDS). Kent, OH: Kent Developmental Metrics.

Reynell, J. K., & Gruber, C., P. (1990). Reynell Developmental Language Scales-U. S. Edition. Los Angeles: Western Psychological Services.

Shepard, L., Kagan, S. L., & Wurtz, E. (Eds.). (1998). Principles and recommendations for early childhood assessments: Submitted to the National Education Goals Panel by the Goal 1 Early Childhood Assessments Resource Group. Washington, D.C.: National Education Goals Panel.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. School Psychology Review, 21, 459-479.

Shinn, M. R., Good, R. H., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. School Psychology Review, 18, 356-370.

Shinn, M. R., III, R. H. G., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. 21st Annual Convention of the National Association of School Psychologists.

Walker, H. M., Severson, H., Stiller, B., Williams, G., Haring, N., Shinn, M., & Todis, B. (1987). Systematic screening of pupils in the elementary age range at risk for behavior disorders: Development and trial testing of a multiple gating model. (pp. 1-57).

Wetherby, A. M., & Prizant, B. M. (1993). Communication and Symbolic Behavior Scales (CSBS). Chicago: Symbolix.

Wetherby, A. M., & Prizant, B. M. (1998). Communication and Symbolic Behavior Scales Developmental Profile. Chicago: Symbolix.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). San Antonio, TX: The Psychological Corporation, Harcourt Brace, Jovanovich.

[1] It should be noted, however, that these key elements may relate, both directly and indirectly, to a traditional task-analytic or critical subskills analysis of a particular general outcome. Indeed, we assume that these links generally will be clear between key elements, as identified for the development of individual growth and development indicators, and critical subskills, as sampled in either Activity-Based Assessment or Ecobehavioral Observations. Please see ECRI-MGD Technical Report 5, *Exploring-Solutions Assessments*, for more information on these latter measures.